

UC Berkeley

UC Berkeley Previously Published Works

Title

GEM-NET: Lessons in Multi-Institution Teamwork Using Collaboration Software.

Permalink

<https://escholarship.org/uc/item/0qc4g77z>

Journal

ACS central science, 5(7)

ISSN

2374-7943

Authors

Gaffney, Stephen G
Ad, Omer
Smaga, Sarah
et al.

Publication Date

2019-07-01

DOI

10.1021/acscentsci.9b00111

Peer reviewed



GEM-NET: Lessons in Multi-Institution Teamwork Using Collaboration Software

Stephen G. Gaffney,[†] Omer Ad,[‡] Sarah Smaga,[‡] Alanna Schepartz,^{*,‡,§} and Jeffrey P. Townsend^{*,†,||}

[†]Department of Biostatistics, Yale University School of Public Health, New Haven, Connecticut 06510, United States

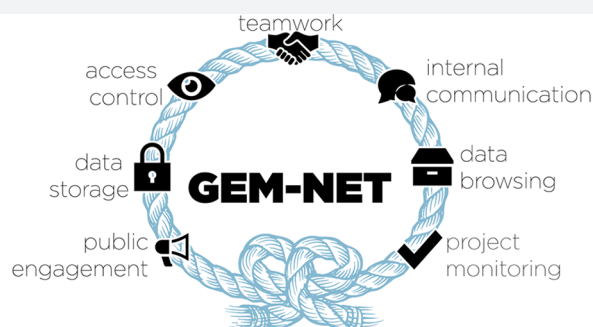
[‡]Department of Chemistry, Yale University, New Haven, Connecticut 06510, United States

[§]Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut 06510, United States

^{||}Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut 06510, United States

S Supporting Information

ABSTRACT: The Center for Genetically Encoded Materials (C-GEM) is an NSF Phase I Center for Chemical Innovation that comprises six laboratories spread across three university campuses. Our success as a multi-institution research team demanded the development of a software infrastructure, GEM-NET, that allows all C-GEM members to work together seamlessly—as though everyone was in the same room. GEM-NET was designed to support both science and communication by integrating task management, scheduling, data sharing, and collaborative document and code editing with frictionless internal and public communication; it also maintains security over data and internal communications. In this Article, we document the design and implementation of GEM-NET: our objectives and motivating goals, how each component contributes to these goals, and the lessons learned throughout development. We also share open source code for several custom applications and document how GEM-NET can benefit users in multiple fields and teams that are both small and large. We anticipate that this knowledge will guide other multi-institution teams, regardless of discipline, to plan their software infrastructure and utilize it as swiftly and smoothly as possible.



Improving the ease and efficiency of collaborative research teams

INTRODUCTION

Multi-institutional teamwork and center-based cross-disciplinary research are associated with high scientific impact^{1,2} and have become ubiquitous within academia^{3–6} and industry.^{7,8} Many academic and industrial teams span multiple and geographically distant research sites, operate in different time zones, and favor different languages. These “virtual” collaborations are supported by a proliferation of Internet-based collaboration tools⁹ that facilitate natural communication between individuals in different locations^{10–12} and across disciplines.^{13–15} The popularity of remote collaboration has prompted efforts to formalize best practices for virtual teams^{16–28} and multiteam systems.²⁹ Nonetheless, it remains challenging to select and customize tools that allow a virtual team within an academic specialty to manage large projects, fulfilling the requirements of task management, scheduling, data sharing,³⁰ data security, document and code collaboration, and scientific and social team communication and coordination.^{31–34} Coordination challenges, in particular, increase with larger and more heterogeneous teams or when more institutions are involved.^{32,35,36} This challenge is more demanding still when constrained by an academic budget. Herein we describe the design and implementation of GEM-NET, a set of more than 20 integrated tools that support the research and broader impact

activities of the Center for Genetically Encoded Materials (C-GEM), an NSF Center for Chemical Innovation. C-GEM spans three universities, at least five disciplines, and six different laboratories and involves research by more than 20 scientists. Here we present our design objectives and motivating goals, provide details of each tool, and discuss the lessons learned during the initial two years of GEM-NET development.

DESIGN OBJECTIVES

A well-designed software infrastructure that adequately supports the interdependent processes of teamwork and taskwork³⁷ can dramatically enhance the success of a collaboration.^{38,39} For C-GEM, we sought to minimize the administrative lag between the conception of research ideas and their execution, promote effective communication and team cognition,^{40,41} develop a knowledge infrastructure that stimulates new ideas, and fully document scientific results.⁴² We also sought to account for the team’s needs in early project phases and anticipate changes in those needs over time, as C-GEM and GEM-NET itself went through iterated phases of planning, implementation, and research translation.⁴³ Our design also considered human and

Received: February 5, 2019

Published: July 12, 2019

Box 1: GEM-NET design objectives

- **Simplicity**
 - A software suite should contain as few tools as possible.
 - Each tool in the suite should be maximally user-friendly to ensure ease and speed of adoption.
 - Each tool in the suite should have a clearly defined purpose, so it's obvious when to use each one.
 - Keeping information up-to-date should require minimal effort and be automated whenever possible.
- **Fast deployment**
 - The majority of tools (in particular communication tools) should provide useful functionality immediately.
 - All team members should be able to use all core tools within weeks.
 - Incremental customization and enhancement of tool functionality is preferred over the complete replacement of temporary tools. This prioritization allows the implementation of software improvements in a manner that minimizes workflow disruptions and allows team members to gradually build proficiency.
- **Easy to maintain**
 - Off-the-shelf, widely adopted, well-supported web applications are preferred over custom-built tools.
- **Better-than-email communication**
 - All communications should be centralized and searchable.
 - The communications interface should provide good separation between topics, and between high-level project discussion and low-level data discussion.
 - It should be easy to add, delete, and restrict the members in any given discussion.
- **Effortless progress tracking**
 - The software suite should provide “at-a-glance” information on project status and progress metrics.
 - Tasks and their associated assignees and deadlines should be clearly organized and easily managed.
- **Manageable notifications**
 - Members should only receive notifications that are personally relevant.
 - Notification frequency should be manageable and customizable to avoid distraction and/or deluge.
- **Organized, portable data**
 - All data should be stored in a secure, easy-to-access, centralized repository.
 - All data should be searchable.
 - Data should be readily exportable from any tool, to ensure full control of data and the possibility of migration to alternative tools if necessary.
 - Services with open APIs should be preferred, to allow data access via scripts, and to enable seamless linking of independent services.
- **Fine-grained access control**
 - There should be a clear and easy-to-manage separation between public and private data.
 - Member laboratories should have complete control over what data they share with collaborators, and shared materials can be read-only.
 - New collaboration members should have immediate access to all data, and access can be discontinued for departing researchers.
- **Tools and data are secure**
 - All members should be able to set their own passwords.
 - Users should be required to memorize as few login details as possible.
- **Reliable access to tools and data**
 - The risks of ‘downtime’ for any tool should be minimized.

social factors^{44,45} and sought to promote a sense of common identity, team spirit, and personal investment in the team's success.⁴⁶ We recognized that in any transdisciplinary collaboration, members inevitably exhibit different degrees of enthusiasm for adopting new platforms and altering working habits.²⁰ The workflows of constituent laboratories may have been refined over many years, so one must be prepared to

accommodate these workflows where necessary or justify the cost in time and effort of significant changes.

■ STEP 1: DEFINING OBJECTIVES

Involving the entire team in software decision-making by consulting everyone about their work patterns, preferred software, and teamwork strategies will result in a well-tailored

Table 1. Pricing, Feature Details, and Setup Notes for the Constituent Tools of GEM-NET^a

TOOL NAME	Benefits	Notable limitations (of free version, where applicable)
Asana	Team task lists	15 user limit per 'Team' within an Organization; 100MB/attachment; no premium features (e.g. Timeline; Portfolios; private projects)
Benchling	Sequence storage, organization, annotation, processing, and search; linked to notebooks to record observations and experiments	10GB storage/user (excl. sequences, oligos); no Inventory; limited tech support; access control on specific-user level only
C-GEM custom tools	Links data across tools, provides team-specific user interfaces	Setup requires some knowledge of web server deployment
└─ Portal	Homepage for team resources, with upcoming calendar items, recent docs, protocols browser and useful links	Incorporating data from APIs of additional tools requires knowledge of Python
└─ Strains	Interactive browser for biospecimen repository; manages shipment requests	Requires a database to keep track of requests, comments, and shipment status
└─ podcast-rss-modify	Customize podcast RSS feed metadata to work with all podcast platforms	Requires limited Python knowledge to modify
git	Version control for code collaboration	Requires familiarity with VCS jargon and command line options
GitHub for Education	Cloud hosting for unlimited version-controlled code repositories and wikis	100GB/repository; 100MB/file; Self-hosted deployment in Enterprise only
└─ GitHub Pages	Free hosting and templates for static web sites through GitHub	100GB/month bandwidth soft limit; 1GB/site; limited plugins supported by Jekyll repositories
Google Scholar	Publication alerts for authors and keywords	Notifications can only be sent to an email address, with no control over alert formatting
GSuite for Education	Bundles user admin and collaboration tools	10k user licenses provided, with more available on request
└─ Team Drive	Shared cloud storage with 'unlimited' storage space	≤400,000 files and folders; ≤5TB for a single file; 750GB/user/day upload before cap activated; cannot share subfolders on a per-user basis (only files or full access)
└─ Google Groups	Access management via email lists	External invitation limit of 500 users/day/group
└─ Google Calendar	Shared calendars	Some limits for extremely heavy usage (e.g. >100k invites in short period)
└─ Google Docs	Collaborative documents	≤ 50 simultaneous editors/commenters; share with ≤ 200 users or groups; 512k characters/doc
└─ Google Sheets	Collaborative spreadsheets	≤ 50 simultaneous editors/commenters; share with ≤ 200 users or groups; 400k cells; 256 columns/sheet; 100 sheets
└─ Google Slides	Collaborative presentations	≤ 50 simultaneous editors/commenters; share with ≤ 200 users or groups; presentations ≤ 10MB
└─ File Stream	Streaming access to Drive files on local machines	Only syncs files within Team Drive
└─ Google APIs	Authentication and data transfers in custom apps	Queries / day limits: 150k for Admin SDK, 1M for Calendar, unlimited for Sheets
IFTTT	Connect apps without resorting to API scripting	Daily usage limits depend on connected services; limited control over frequency of updates
Jekyll	Easily generate static websites with Markdown-based blogging	For static sites only
Mendeley	Shared reference library: sync publications with website	2GB storage/user ; ≤3 members per private group created by free account
Paperpile	Add citations/bibliography to Google Docs via free add-on	Organize and import reference library in premium only
Slack	Instant, searchable messaging organized in topic 'channels', with numerous 3rd party integrations	Free version limits message history to most recent 10k messages, and lacks 'user groups' for group-level channel membership
Soundcloud	Podcast hosting, playback stats, embeddable streaming player	≤3h upload time in free version; extended/unlimited uploads in premium versions
Twitter	Communicate with worldwide network of scientists and the public	Messages have 280 character limit; accounts have a single associated email address and password
└─ TweetDeck	Designate multiple users to tweet for team without needing to share account password	Only works with TweetDeck browser interface; no way to trace which user tweeted
Zoom	Video conferencing with recording, simultaneous screen sharing, remote control, and join by phone	≤40min meetings and ≤100 participants; premium versions have extended durations, and more participants, admin controls

^aSupplementary Table 1 provides additional information about costs, setup dependencies, and scaling for large teams.

design in which the team can feel invested.⁴⁷ In addition to representing a wise starting point for software infrastructure design, an early consultation process allows a smooth transition to the vital cycle of seeking feedback⁴⁸ on user experience and responding with refinements that will continue for the life of the collaboration. Once informed by discussions with the team, the

infrastructure designer should create a list of objectives that define the characteristics and functionality of the collaboration tools. These objectives will serve as a checklist against which software options can be evaluated and a reference point for future software changes or customizations. This process is the equivalent of writing the all-important “functional specification”

in traditional software design.⁴⁹ Pinning down objectives before launching into implementation gives the best chance for an optimized design and swift construction. The objectives list for GEM-NET (Box 1) is likely to be widely applicable across fields and team sizes.

TOOLS

On the basis of our design objectives, we assembled a suite of over 20 tools, including two custom-built applications that fulfilled the needs of C-GEM. All tools are free or use the “freemium” model (offering premium features for a fee). For each tool, we have gathered details of benefits, version limitations, costs, setup times, and dependencies and information on how they scale to large teams (Table 1, Supplementary Table 1). The tools are highly interlinked (Figure 1) but can be broadly partitioned into seven categories:

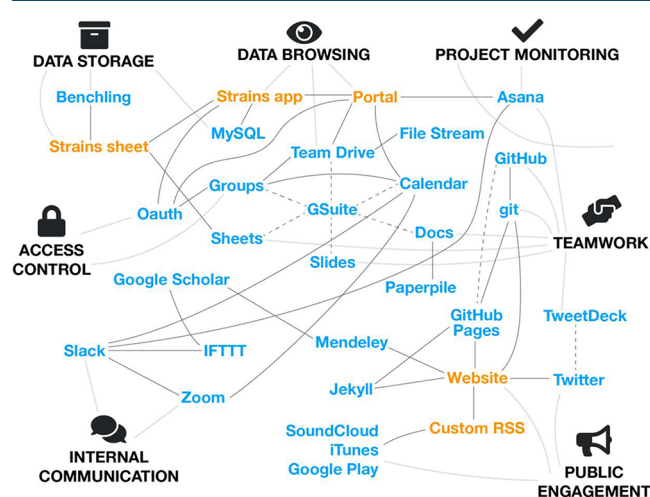


Figure 1. GEM-NET toolset. Connections indicate tool categorizations and data flow between services. Dashed lines indicate platform and subservice relationships. Custom applications are shown in orange, and off-the-shelf tools are shown in blue.

access control, data storage, data browsing, project monitoring, teamwork, internal communication, and public engagement. These categories are described in turn later.

USER ADMINISTRATION AND ACCESS CONTROL (G SUITE FOR EDUCATION, GOOGLE GROUPS)

Associating each team member with a primary online identity and using this identity for tool access greatly simplifies access control and links team members to their associated data across services. To manage these identities, and to satisfy our requirements for fine-grained access control, we took advantage of Google’s Web-based applications and services⁵⁰ through Google’s G Suite. G Suite can create and manage user accounts for an organization and bundle a wide range of Google’s popular services (Google Calendar, Docs, Sheets, Drive, and many more) that are linked to the organization’s domain name. As an official university-affiliated center with responsibilities for teaching and training graduate students and postdoctoral associates, C-GEM was eligible for the free academic edition, G Suite for Education, which we linked to our domain “gem-net.net”.

For maximum flexibility in setting up online identities for the team and managing access privileges, we use Google Groups.

Google Groups is a service that manages e-mail address lists that can be used to access control and mailing lists. Every C-GEM member was added to a user group called everyone@gem-net.net using a domain-associated account (e.g., “alanna@gem-net.net”) or generic Google-associated account of their choice. Permissions assigned to a user group are inherited by everyone in that group, so adding the “everyone” group to Calendar and Team Drive, for example, grants everyone access to these services. This association of user groups with authorization allows a “single sign-on” approach, where one needs only a single set of credentials to access multiple services, and readily extends to custom apps. Custom apps can use Google’s OAuth service for authentication and test for group membership using the Directory API to unlock private features. We found that all members opted for the convenience of using a pre-existing account rather than creating a new “gem-net” account, allowing C-GEM resources to be immediately visible in Google applications without the burden of account switching.

Group-based authorization can also be used to provide flexibility in account deprovisioning. Departing researchers could be moved from a full-access group to a partial access-group, granting them read-only access to select data and applications. For fine control over permissions, groups can be nested within other groups, inheriting their access.

DATA STORAGE (TEAM DRIVE, BENCHLING, STRAINS LIST)

To tailor a data storage strategy to the specific data types being handled, tool selection should incorporate the consideration of likely file sizes, formats, indexing requirements (which features of the data should be searchable?), and editing requirements (will files be modified after storage?). C-GEM’s data needs led us to the combination of a general purpose tool, a tool geared toward sequence data, and a custom tool to serve as a directory for our biospecimen repository.

Our general purpose file storage tool is Google’s Team Drive. Team Drive is the cloud file management service in G Suite, extending the functionality of the personal-account Drive service by providing a shared file space that can be accessed by a whole team. At the time of writing, the Team Drive provides unlimited file space and allows members to upload and sync files up to 5 TB in size, providing an opportunity for far more than archival storage of drafts and publications.⁵¹ C-GEM uses this service to share all of its documents, including meeting minutes, progress reports, proposals, protocols, blog post drafts, all organized in subdirectories that can be accessed in Web browsers, in the Google Drive mobile app, or on the desktop using the File Stream app (note: File Stream requires a domain-associated user account).

For the purpose of storing and managing nucleotide sequence data, the raw data used frequently by C-GEM, we use the sequence-focused platform Benchling. Benchling allows sequences to be manually imported, created from scratch, or directly downloaded from various public repositories using a search interface. Sequences are grouped into “projects” and stored in either a personal account or in the common storage area of an “Organization” account. Our team stores its sequences in the “GEM-NET” organization or in lab-specific organizations as appropriate. Benchling includes an array of tools for annotating, analyzing, aligning, and manipulating sequences, providing URLs that let sequences be shared in editable or read-only form. Benchling gives each sequence a unique identifier, which

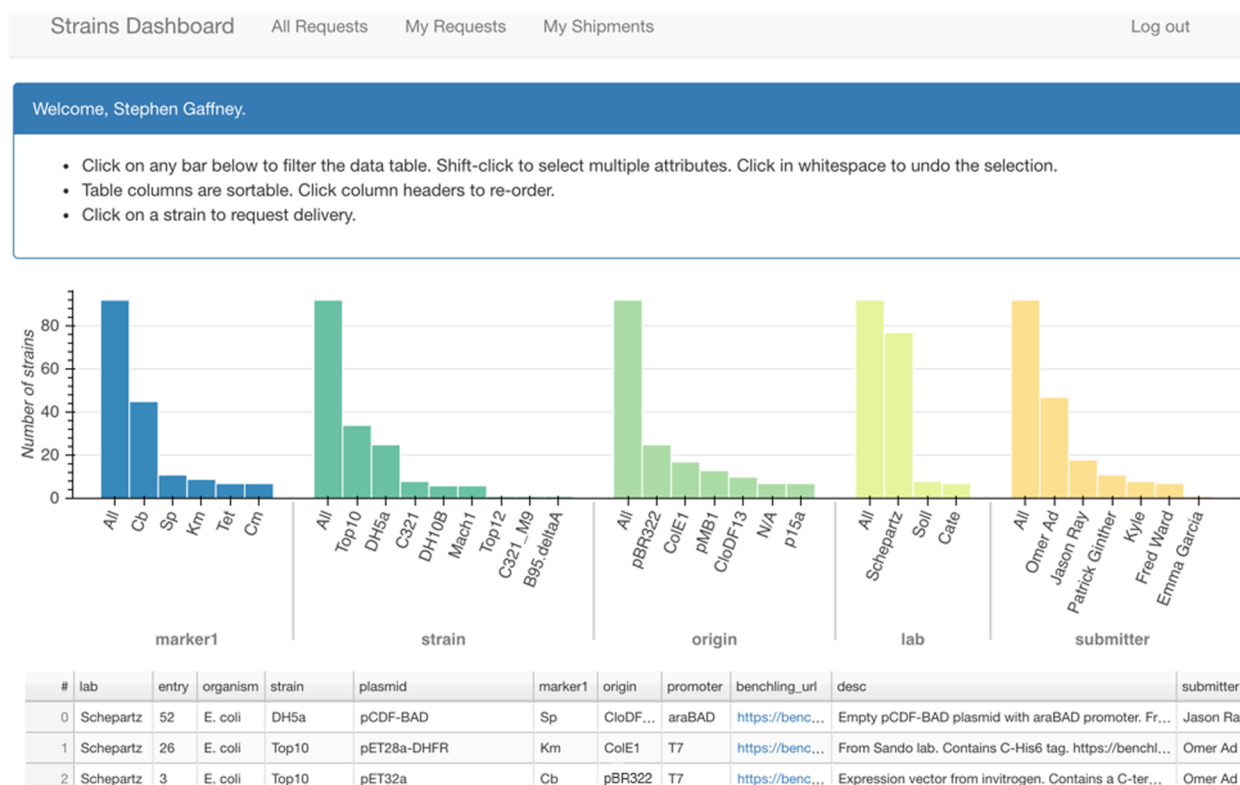


Figure 2. Homepage of the C-GEM Strains app, a custom application used to interact with the C-GEM biospecimen repository. Users can browse the repository, request samples, and track the shipment of sample requests. Interactive bar charts and a sortable table aid in filtering the sample collection.

allows C-GEM to link every experimental result back to the original sequence and its associated metadata.

The third strand of our data storage strategy is the management of our biospecimen repository, also known as the C-GEM Strains List. The Strains List contains the subset of Benchling sequences that correspond to “real-world” physical specimens in our laboratories. This repository is stored as a shared spreadsheet, built using Google Sheets and accessible in the Team Drive. The fields included in the spreadsheet were selected so that each strain record is clear and unambiguous but minimal enough that it is not laborious to manually add new entries. Fields include physical location, organism (e.g., *E. coli*), strain (e.g., *DH5alpha*), plasmid name, marker (e.g., *Cb*), origin (e.g., pBR322), promoter (e.g., T7), extended description, benchling URL, and submitting member. By gathering and maintaining this information as a shared resource, we ensure its continuity when team members depart and allow current and future members to know what is in our specimen collections at a glance. We intend to release a version of this repository to the public, as its size and meticulous annotation would make it a valuable resource for the biochemistry community.

■ DATA BROWSING (STRAINS APP, PORTAL, PROTOCOLS)

Whereas tools for data storage generally include the ability to browse and search the stored data, achieving the desired level of data querying functionality may require the construction of additional purpose-built tools. Creating a new layer of software for browsing and searching has the additional benefits of enabling data aggregation from multiple independent tools and helping to maintain data integrity, preventing accidental modification by separating the data entry from data reading

operations. Building these custom applications requires that the associated data storage tools expose their data through open APIs. Using these APIs and constructing custom interfaces requires programming knowledge, but this area is one where a little effort goes a long way: most of the development “heavy lifting” is done by libraries that allow one to create a basic Web server with only a few lines of code, and even the most simple data merging functionality can pay dividends. C-GEM uses two custom applications that harness tool APIs and shares all source code for both on its GitHub account (<https://github.com/gem-net>). Both tools were built by a single developer (the first author of this Article) in the Python language using the Flask Web server library.

Our first purpose-built tool is the C-GEM Strains App. This application has two main functions: It offers a convenient way to browse the strain repository, and it offers “shipment requests”, allowing any member of the team to request the delivery of any given strain. Leveraging the Google Sheets API, the Strains App gathers the strains from each lab worksheet and uses the Bokeh visualization library to create an interactive browsing interface (Figure 2). Strain characteristics (e.g., the *marker*, *strain*, and *origin* fields) are aggregated in a bar chart, and selecting these bars (by touch or mouse-click) whittles down the data, which are shown in a sortable table underneath. Clicking on a strain in the data table brings the user to a request form to order the strain and provide contact details. The appropriate lab is notified of new requests (with notifications delivered to the members listed in the Strains List) and can discuss shipment progress with the requester on a dedicated discussion page. Overview pages allow all shipment requests to be monitored at a glance.

Our second custom application is the GEM-NET Portal, a private resource for C-GEM members. The portal provides a

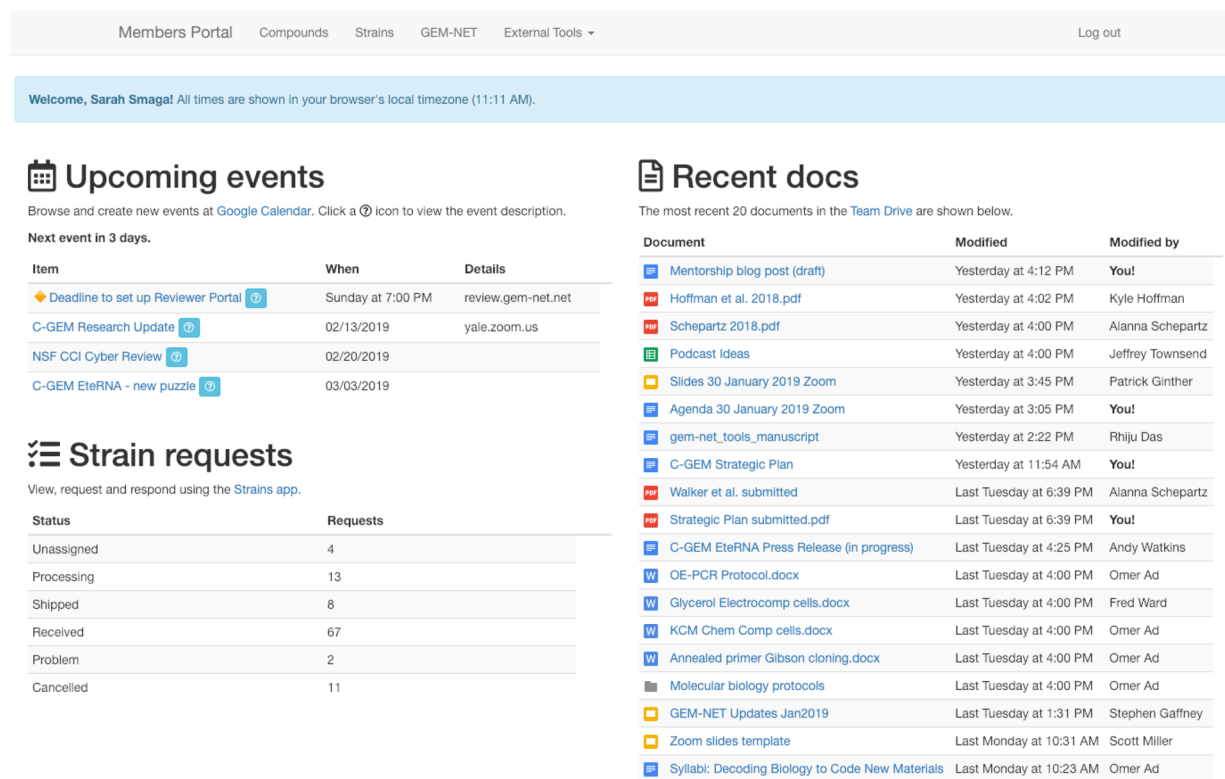


Figure 3. Homepage of the GEM-NET Portal. The homepage gathers data from multiple GEM-NET tools to provide a unified interface for team resources. This version includes upcoming events, aggregated strains shipment data, and recent documents from the Team Drive.

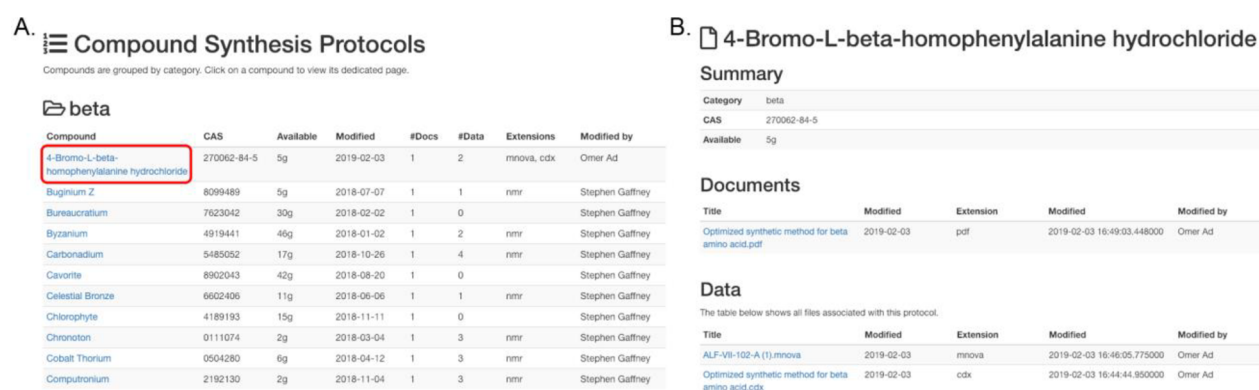


Figure 4. GEM-NET Portal also stores synthesis protocols and characterization data for C-GEM compounds. (A) Available information is summarized in the landing page. (B) Hyperlinks bring users to compound-specific pages.

handy dashboard of resource links, upcoming Google Calendar items, recently modified documents in our Team Drive, and a summary of active and completed shipment requests from our Strains app and will soon be expanded to show personalized task lists (Figure 3). By pulling data from all of the C-GEM tools, the portal provides a high-level summary of data and progress in a single location, reducing the burden on members to log in to individual services.

The portal is also a natural home for resources that team members need to access frequently. Where these resources are files or folders in the Team Drive, their metadata can be extracted using the Google Drive API and used in specialized browsing interfaces within the Portal. Indeed, this approach is currently being used for our growing collection of chemistry protocols. The experimental protocol for each synthetic step used to produce a desired compound is stored in its own

subdirectory of our “chemistry protocols” Team Drive folder, comprising a document and corresponding image and data files (e.g., NMR). To make these folders as easy to browse as possible, the Portal features a dedicated Protocols page where entries can be explored at a user level without having to open Google Drive (Figure 4).

PROJECT MONITORING

Our project monitoring needs are met by Asana. Asana is a project management tool that lets teams organize and follow tasks and subtasks in lists or kanban-style boards, designate assignees and due dates, and attach files and comments to task items. C-GEM structures its task lists as six Asana “Projects” that correspond to the six “Impact Areas” delineated during our initial Strategic Planning Meeting: Management, Education, Innovation, Participation, Communication, and Research. Using

the Asana API, we imported our lists of action items for each project, setting assignees, due dates, and extended descriptions. With initial tasks in place, each member could see an “inbox” list of tasks assigned to them (or any other team member) ordered by due date and would receive weekly summaries and notifications of impending deadlines. For highly interdependent tasks, sublists of items were created and assigned to different individuals. This centralized “to-do list” allows everyone on the team to monitor the progress of the collaboration in real time and provides a record of this progress through the history of completed tasks.

■ TEAMWORK: COLLABORATING ON DOCUMENTS AND CODE

The ability to work together on documents is crucial to any academic collaboration, and for this task we use Google Docs. Google Docs has transformed the way people can remotely collaborate on documents by allowing multiple users to work on the same document simultaneously.⁵² This simultaneous access removes the need to circulate draft documents by e-mail, accelerating the writing and revision process—an important consideration for an agile team with trainees as major contributors who have an urgent need to publish⁵³—and also provides a way to designate access privileges to contributing users. We also make use of Google Sheets and Google Slides, which offer the same concurrent editing ability for spreadsheets and presentations, respectively. Each of these cloud services maintains a history of all file changes, with the ability to create named versions, making it easy to revert to past versions. They also make it possible to create comment threads on any part of the file and convert comments into tasks that can be assigned to specific users.

When writing manuscripts, our team uses the Paperpile Google Docs add-on to add citations. The free version of Paperpile allows one to search for publications based on identifiers such as author name, title, DOI, or ISBN, to edit reference details, and to generate citations and document bibliography using journal-specific formatting, all without leaving the document.

For collaborative reference management outside of Google Docs, we use Mendeley. Mendeley lets teams create public or private groups that store a shared reference library, accessible by a Web browser or a desktop app. Browser plugins and search functionality make it very easy to add references to the library, and references can be exported in multiple formats. C-GEM uses Mendeley to manage its Web site publication list.

Our team also collaborates on code. C-GEM code is organized in “repositories” on GitHub using a free GitHub for Education organization account. GitHub repositories can be synchronized with local copies on team members’ machines, where they are managed with the tool “Git”. Git is a version control system that maintains a complete version history of the code and allows features to be developed in isolated “branches” before being merged with the main codebase. The C-GEM Web site,⁵⁴ the Strains App, and the GEM-NET Portal are all developed using Git and GitHub.

■ INTERNAL COMMUNICATION AND COORDINATION

C-GEM has managed to almost entirely replace e-mail using the communication software Slack. Slack is a messaging system that can be used from a Web browser or from mobile and desktop

apps. Messages are published to topic-specific “channels” which can either be public to the team or private to a list of members. Slack also offers private, direct messaging, and fine-grained control over user notifications. This centralization of our team communication makes it very easy to communicate, find messages when they are needed, and share message history with new members. Slack has become our go-to platform for day-to-day scientific interaction, such as protocol sharing and troubleshooting. Slack features emoji, which have been demonstrated to enhance user satisfaction and convey a sense of personal communication.^{55,56}

Slack can be integrated with a large number of third party services that can enhance its functionality. Using the Asana integration, our team members are notified in a private channel when tasks are assigned to them or when a deadline is approaching, and they can update the tasks directly from Slack. To stay up to date with current literature, we created a gem-net domain user account that subscribes to multiple Google Scholar alerts, and use an IFTTT applet to generate new posts in a “publications” channel when alert e-mails arrive.

C-GEM meetings are primarily conducted as video conferences using the software Zoom. We have access to Pro accounts through Yale University, which (as with free accounts) can support up to 100 participants in a call. To maintain a record of our meetings, we record meeting minutes and comments live in a shared Google Docs file and add slides to a shared Google Slides presentation. Presenters are granted control of the meeting slides using Zoom’s remote control function.

To schedule meetings and other events, C-GEM uses a centralized calendar in Google Calendar. As with other Google tools, we control access based on the membership of our “everyone” Google Group, allowing team members to see the C-GEM calendar alongside their personal calendars. Google Calendar allows timezone-specific scheduling, which is useful for geographically dispersed collaborations, and its API and Slack integration make it easy to link to other applications.

■ PUBLIC ENGAGEMENT

To share news and events with the outside world, we use three primary outlets: Twitter, the GEM-NET Web site, and the C-GEM podcast.

Our team makes frequent use of Twitter using the handle @cgem_cci to share short updates, to keep up with breaking science news, and to converse directly with the public. Propagating scientific knowledge via Twitter has been argued to broaden audiences, speed up communication, force clearer and more concise descriptions of a researcher’s work,⁵⁷ and even potentially reorganize semantic thinking.⁵⁸ In August 2018, we held our first Twitter Q&A in partnership with the American Chemical Society, responding to questions from the public that included the hashtag “#AskCGEM”. To allow multiple team members to tweet from our account, we use the “Team” feature of TweetDeck, specifying personal accounts that can post on behalf of C-GEM.

The GEM-NET Web site (<http://gem-net.net>) is the central resource for information about C-GEM. It details our mission and gives a primer on the underlying science; it introduces our team, with bios for investigators; and it includes our publications list, contact information, and links to other C-GEM resources on the Web. Frequently updated information can be found on our “Updates” page, where visitors can see our Twitter feed and listings for C-GEM news and events. The Web site is also home to our “Chemical Matters” blog, where our team posts about

recent advances in the literature, and features a landing page for our podcast, with the ability to stream episodes and read show notes for each episode. The Web site is set up to be easily modified to accommodate new types of content in the future, as we develop educational materials, videos, and a partnership with the citizen-science videogame EteRNA.⁵⁹

Our Web site was built using the static Web site generator Jekyll. Jekyll simplifies Web site building by separating text content from code for page structure and styling and can generate a blog (or other collection of pages with a shared template) from a folder of text files. Text content can be written in Markdown, which provides an intuitive, easy-to-learn, easy-to-read syntax for creating Web page elements that Jekyll converts to HTML. This ease-of-use enables anyone on our team to update Web site content. Because Markdown is a superset of HTML, we can mix in HTML to accomplish functionality that is not covered by the basic syntax.

There are numerous prebuilt Jekyll “themes” available online that can be used to handle page structure and styling. We use the blog-oriented “Minimal Mistakes” theme,⁶⁰ which allows public commenting on blog posts, among other features. Plugins enable more complicated functionality, such as generating the RSS feed for our blog (using jekyll-feed), automatically building our interactive publications page from a Mendeley bibtext file (with jekyll-scholar), and embedding our Twitter feed (with jekyll-twitter-plugin). When performing Web site updates, we first test changes locally using Jekyll’s built-in Web server, then push the final Web site files to GitHub, which provides free Web site hosting through GitHub Pages.

Our other public outlet is our podcast, which we record and release with the help of the Yale Broadcasting and Media Center. There are numerous podcasting hosting platforms available. We opted for SoundCloud, releasing our episodes as a “set” (i.e., playlist) within the Yale’s SoundCloud account. For each set, SoundCloud generates an RSS feed, the file used by podcast aggregation and subscription services. This feed is compatible with iTunes, one of the largest podcast aggregators. Uploading the feed to iTunes makes the podcast available through Apple Podcasts and numerous other subscription services that automatically pull from the iTunes Store, such as Overcast, PlayerFM, and Blubrry. Some aggregators require extra information to be embedded within the RSS feed, so to maximize compatibility (and further customize metadata, such as images and contact information), we modify our SoundCloud feed using a script that is available on our GitHub account.⁶¹ We host this customized feed on our Web site and use it to list the podcast with Google Play Music, TuneIn (which makes the podcast available through Amazon’s Alexa), Stitcher, Spotify, and others, all of which provide listings for free.

■ DISCUSSION: LESSONS FROM GEM-NET

In the course of developing our current software infrastructure, we encountered the occasional instructive challenge—little quirks in software or bumps in the road to teamwork efficiency. We list a few of these here, in the form of six lessons.

Lesson 1: *There are no perfect all-in-one tools.* Our initial efforts at software infrastructure design focused on tools that bridged categories, meeting as many of our needs as possible, but these invariably lost out to tools that do one thing well. We hoped, for example, to find software that can manage tasks and handle all internal communication. But the messaging options provided by task management software were too limited or cumbersome to suit the flexibility and control we desired, and the messaging-

oriented tools we tried lacked the full functionality we hoped for in task management software. This example is one of several cases of overlap in tool functionality, but our setup avoids confusing redundancy by matching a single purpose with a single tool (satisfying our “simplicity” objective). A desire to limit costs also limits potential redundancies, as the free versions of the chosen software have less overlap. For instance, Paperpile could replicate Mendeley’s reference library management, but only in the premium version, whereas Mendeley cannot format citations in a Google Doc. Slack could support multiuser video conferencing, but only in the premium version. Our choice of Zoom for video conferencing rather than Hangouts Meet, which is included in G Suite, was based on its recording and remote control abilities and—crucially—on its service being provided free through our university. Hangouts Meet could be a good free alternative for other teams.

Lesson 2: *Trial new tools with real tasks, not generic demos.* We have had the most success launching new tools when team members are given a walkthrough that performs a real, team-relevant task, shortly followed by the team trying out the tool to accomplish something that actually needs to be done. This boosts learning, promotes software uptake, quickly reveals shortcomings of the tools, and makes efficient use of everyone’s time. A corollary is that tools are ready for launch only after enough customization and testing to handle real use.

Lesson 3: *Some tools are only effective once everyone is using them.* This is especially true for messaging and task management. If the whole team is not active on Slack, then your message to the team will not get to everyone, and you will find yourself falling back on e-mail. If members do not update tasks in Asana, project statuses quickly diverge from reality. Both of these scenarios happened in the early days of C-GEM. We found e-mail particularly hard to phase out and still fall back on it occasionally. The tools listed in this paper are still *useful* to those who use them, but that usefulness is limited, and partial uptake may lead to inefficient, redundant systems, and faltering usage. The solution is in the next two lessons.

Lesson 4: *A team leader should make tool use mandatory.* The quickest way to achieve full uptake for a tool is to have someone in a leadership role apply pressure on everyone to use it. If the tool is worthwhile, then the utility of the tool (and cost of not using it) will be evident to the team in the first use, so an initial nudge should suffice to solve the coordination problem of Lesson 3.

Lesson 5: *Administrative oversight and role designation are important.* In principle, the software listed here could let a team operate in a self-organizing way, with everyone adding their own tasks, posting blog entries and tweets, and keeping data up to date. However, in practice, designated administrators will prove necessary for chasing things up and filling in gaps. We rely on an administrator to take the lead writing up minutes, to create and assign tasks based on meeting discussions, to check up on people near deadlines, to send reminders about upcoming meetings, and to nudge people to update collaborative documents. Building momentum on Twitter also required a designated principal tweeter, and updating the Web site fell to a specific team member. (Markdown proved a bridge too far for most of the team, so blog posts are often submitted as Word documents, then converted by the Web site administrator.) Even more importantly, the administrator is responsible for monitoring and promoting team engagement on GEM-NET and identifying any barriers that may be hindering participation.

Lesson 6: *Make the occasional concession to chaos (or, perfect is the enemy of good)*. In some cases, we found that a planned “optimum” workflow gave way to an easier-to-attain alternative, as with the blog posting method mentioned above. Another example is in Benchling, where sequence data would ideally be organized within projects in the C-GEM Benchling account and annotated according to a precise set of rules. This would obviate the need for an independent Strains List, make all sequence data available through the Benchling API, and allow our Strains app to extract the repository directly from Benchling and offer extended functionality. Instead, we found it more practical for laboratories and individual users to manage their sequences however it suited them. We also found that the simplicity and ease of access of the independent Strains List made it very easy to keep up to date and helped it become one of our most frequently used tools—benefits that may outweigh the potential data redundancy.

CONCLUSIONS

The activities of all collaborative research teams, regardless of discipline or size, are well-served by applications that address the seven key areas of access control, data storage, data browsing, project monitoring, teamwork, internal communication, and public engagement. A universal challenge is to identify a minimal set of applications that fully satisfies the needs of a team in each of these areas while remaining both user- and budget-friendly. GEM-NET, the set of tools that make up the software infrastructure of the NSF Center for Genetically Encoded Materials (C-GEM), offers a solution to this challenge that was stress-tested over two years by all team members. GEM-NET supported and enhanced C-GEM’s collaborative research goals by organizing, documenting, and simplifying the communication of ideas, data, and analysis. Assembled according to a carefully considered set of design objectives (Box 1) and fine-tuned according to team feedback, GEM-NET meets the wide-ranging collaboration needs of the team while remaining intuitive in structure, so that every team member knows what tool to use for any particular task and simultaneously provides a high level of security, reliability, and ease of maintenance. Some tools are off-the-shelf and others are novel, but all are both transferable and scalable.

Our aim has been to provide sufficient details of the function of each tool within GEM-NET and how these tools connect to and complement each other to help other teams easily reproduce our setup or adapt it to their needs. We provide information on feature distinctions between free and premium tool versions (all of the tools use the free or freemium model), upgrade pricing (to support budgeting for the costs of increasing team size or requirements), and pointers on how to adapt these tools to work with large teams, including how the tools support work within subteams (Supplementary Table 1). Notably, the vast majority of tools within GEM-NET are agnostic of research area and thus are suitable for any collaborative team. For teams that handle sequence data, GEM-NET can be replicated in its entirety. Other teams may wish to swap Benchling, our chosen sequence management tool, for an alternative platform geared toward another data format or to manage all data through Team Drive; its unlimited storage and comprehensive API work with any data type. Our custom apps demonstrate how simple, tailor-made code can harness the APIs of powerful, well-established software to merge data between services and create new functionality. We have documented and shared the source code of our apps, including a biospecimen repository browser with

interlab shipment tracking and a members portal with protocols listing.

Because most of the functionality works “out of the box”, teams can use GEM-NET to establish a collaborative network with minimal delay: jointly authoring documents and code, sharing data, video conferencing, scheduling meetings, tracking progress, and publicizing their work through a Web site, blog, podcast, and Twitter feed. Supplementing the initial tool set with custom applications that merge data across platforms, we gained additional abilities to browse and query our data and produce a unified interface through a member portal. The first steps in reproducing GEM-NET are to register a domain, to set up a basic Web site, and then to apply for a G Suite for Education license. Once the license is granted, set up admin user accounts and then user groups in Google Groups, using these to provision access to the range of G Suite tools. Next, set up organizations in Benchling, Asana, GitHub, and Slack. The remaining tools can be set up in any order, satisfying the dependencies listed in Supplementary Table 1.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acscentsci.9b00111](https://doi.org/10.1021/acscentsci.9b00111).

Expanded Table 1 including information about cost, setup dependencies, and scaling tools for larger teams (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: alanna.schepartz@yale.edu.

*E-mail: jeffrey.townsend@yale.edu.

ORCID

Sarah Smaga: 0000-0003-3469-2642

Alanna Schepartz: 0000-0003-2127-3932

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Center for Genetically Encoded Materials, an NSF Center for Chemical Innovation (NSF CHE-1740549). O.A. was supported in part by Agilent Technologies as an Agilent Fellow. The authors are also grateful to Zhang Zhang for insightful suggestions during the preparation of this manuscript.

REFERENCES

- (1) Sanfilippo, P.; Hewitt, A. W.; Mackey, D. A. Plurality in Multi-Disciplinary Research: Multiple Institutional Affiliations Are Associated with Increased Citations. *PeerJ* **2018**, 6, No. e5664.
- (2) Hall, K. L.; Stokols, D.; Stipelman, B. A.; Vogel, A. L.; Feng, A.; Masimore, B.; Morgan, G.; Moser, R. P.; Marcus, S. E.; Berrigan, D. Assessing the Value of Team Science: A Study Comparing Center- and Investigator-Initiated Grants. *Am. J. Prev. Med.* **2012**, 42 (2), 157–163.
- (3) Jones, B. F.; Wuchty, S.; Uzzi, B. Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science. *Science* **2008**, 322 (5905), 1259–1262.
- (4) Sienkiewicz, J.; Soja, K.; Holyst, J. A.; Slood, P. M. A. Categorical and Geographical Separation in Science. *Sci. Rep.* **2018**, 8 (1), 8253.
- (5) Adams, J. D.; Black, G. C.; Clemmons, J. R.; Stephan, P. E. Scientific Teams and Institutional Collaborations: Evidence from U.S. Universities, 1981–1999. *Res. Policy* **2005**, 34 (3), 259–285.

- (6) Wuchty, S.; Jones, B. F.; Uzzi, B. The Increasing Dominance of Teams in Production of Knowledge. *Science* **2007**, 316 (5827), 1036–1039.
- (7) Hagedoorn, J. Inter-Firm R&D Partnerships: An Overview of Major Trends and Patterns since 1960. *Res. Policy* **2002**, 31 (4), 477–492.
- (8) Matt, L. N. H. Partnering for Progress: How Collaborations Are Fueling Biomedical Advances. *Drug Dev. Delivery* **2017**, 17 (8), 20–24.
- (9) Orta-Castañón, P.; Urbina-Coronado, P.; Ahuett-Garza, H.; Hernández-de-Menéndez, M.; Morales-Menéndez, R. Social Collaboration Software for Virtual Teams: Case Studies. *Int. J. Interact. Des. Manuf.* **2018**, 12 (1), 15–24.
- (10) Haman, M.; Hertzum, M. Collaboration in a Distributed Research Program. *J. Doc.* **2019**, 75 (2), 334–348.
- (11) Battke, F.; Symons, S.; Herbig, A.; Nieselt, K. GaggelBridge: Collaborative Data Analysis. *Bioinformatics* **2011**, 27 (18), 2612–2613.
- (12) Lasater, K.; Johnson, E.; Hodson-Carlton, K.; Siktberg, L.; Sideras, S. A Digital Toolkit to Implement and Manage a Multisite Study. *J. Nurs. Educ.* **2012**, 51 (3), 127–132.
- (13) Lee, Y.-L.; Chien, T.-F.; Hsu, S.-C.; Chang, P. Developing and Applying a Cross-Disciplinary Team Handover Information System. *Stud. Health Technol. Inform.* **2013**, 192, 1185.
- (14) Lee, E. S.; McDonald, D. W.; Anderson, N.; Tarczy-Hornoch, P. Incorporating Collaboratory Concepts into Informatics in Support of Translational Interdisciplinary Biomedical Research. *Int. J. Med. Inform.* **2009**, 78 (1), 10–21.
- (15) Calhoun, W. J.; Wooten, K.; Bhavnani, S.; Anderson, K. E.; Freeman, J.; Brasier, A. R. The CTSA as an Exemplar Framework for Developing Multidisciplinary Translational Teams. *Clin. Transl. Sci.* **2013**, 6 (1), 60–71.
- (16) Stokols, D.; Misra, S.; Moser, R. P.; Hall, K. L.; Taylor, B. K. The Ecology of Team Science: Understanding Contextual Influences on Transdisciplinary Collaboration. *Am. J. Prev. Med.* **2008**, 35 (2 Suppl), S96–S115.
- (17) Salmons, J. *Handbook of Research on Electronic Collaboration and Organizational Synergy*; IGI Global, 2008.
- (18) Murugesan, S. *Handbook of Research on Web 2.0, 3.0, and X.O: Technologies, Business, and Social Applications: Technologies, Business, and Social Applications*; IGI Global, 2009.
- (19) Kock, N. *E-Collaboration: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*; IGI Global, 2009.
- (20) Nunamaker, J. F., Jr.; Reinig, B. A.; Briggs, R. O. Principles for Effective Virtual Teamwork. *Commun. ACM* **2009**, 52 (4), 113.
- (21) Salas, E.; Cooke, N. J.; Gorman, J. C. The Science of Team Performance: Progress and the Need for More. *Hum. Factors* **2010**, 52 (2), 344–346.
- (22) Salas, E.; Reyes, D. L.; McDaniel, S. H. The Science of Teamwork: Progress, Reflections, and the Road Ahead. *Am. Psychol.* **2018**, 73 (4), 593–600.
- (23) National Research Council, Division of Behavioral and Social Sciences and Education; Board on Behavioral, Cognitive, and Sensory Sciences; Committee on the Science of Team Science. *Enhancing the Effectiveness of Team Science*; National Academies Press, 2015.
- (24) Gostick, A.; Elton, C. *The Best Team Wins: The New Science of High Performance*; Simon and Schuster, 2018.
- (25) Cummings, J. N.; Kiesler, S. Organization Theory and the Changing Nature of Science. *J. Org. Des.* **2014**, 3 (3), 1–16.
- (26) Kozlowski, S. W. J. Enhancing the Effectiveness of Work Groups and Teams: A Reflection. *Perspect. Psychol. Sci.* **2018**, 13 (2), 205–212.
- (27) Kozlowski, S. W. J.; Ilgen, D. R. Enhancing the Effectiveness of Work Groups and Teams. *Psychol. Sci. Public Interest* **2006**, 7 (3), 77–124.
- (28) Bell, B. S.; Kozlowski, S. W. J. A Typology of Virtual Teams. *Group & Organization Management* **2002**, 27 (1), 14–49.
- (29) Shuffler, M. L.; Carter, D. R. Teamwork Situated in Multiteam Systems: Key Lessons Learned and Future Opportunities. *Am. Psychol.* **2018**, 73 (4), 390–406.
- (30) Tan, A.; Tripp, B.; Daley, D. BRISK—Research-Oriented Storage Kit for Biology-Related Data. *Bioinformatics* **2011**, 27 (17), 2422–2425.
- (31) Freeman, R. B.; Ganguli, I.; Murciano-Goroff, R. Why and Wherefore of Increased Scientific Collaboration. In *The Changing Frontier*; Jaffe, A. B., Jones, B. F., Eds.; University of Chicago Press, 2014; pp 17–48.
- (32) Cummings, J. N.; Kiesler, S. Coordination Costs and Project Outcomes in Multi-University Collaborations. *Res. Policy* **2007**, 36 (10), 1620–1634.
- (33) Hall, W.; De Roure, D.; Shadbolt, N. The Evolution of the Web and Implications for eResearch. *Philos. Trans. R. Soc., A* **2009**, 367 (1890), 991–1001.
- (34) Zhang, Z.; Cheung, K.-H.; Townsend, J. P. Bringing Web 2.0 to Bioinformatics. *Brief. Bioinform.* **2008**, 10 (1), 1–10.
- (35) Cummings, J. N.; Kiesler, S.; Bosagh Zadeh, R.; Balakrishnan, A. D. Group Heterogeneity Increases the Risks of Large Group Size: A Longitudinal Study of Productivity in Research Groups. *Psychol. Sci.* **2013**, 24 (6), 880–890.
- (36) Cummings, J. N.; Kiesler, S. Collaborative Research Across Disciplinary and Organizational Boundaries. *Soc. Stud. Sci.* **2005**, 35 (5), 703–722.
- (37) Fiore, S. M.; Wiltshire, T. J. Technology as Teammate: Examining the Role of External Cognition in Support of Team Cognitive Processes. *Front. Psychol.* **2016**, 7, 1531.
- (38) Easley, R. F.; Devaraj, S.; Michael Crant, J. Relating Collaborative Technology Use to Teamwork Quality and Performance: An Empirical Analysis. *J. Manage. Inf. Syst.* **2003**, 19 (4), 247–265.
- (39) Olson, J. S.; Olson, G. M. How to Make Distance Work Work. *Interactions* **2014**, 21 (2), 28–35.
- (40) DeChurch, L. A.; Mesmer-Magnus, J. R. The Cognitive Underpinnings of Effective Teamwork: A Meta-Analysis. *J. Appl. Psychol.* **2010**, 95 (1), 32–53.
- (41) Salas, E.; Fiore, S. M.; Letsky, M. P. *Theories of Team Cognition: Cross-Disciplinary Perspectives*; Routledge, 2013.
- (42) Olson, J. S.; Olson, G. M. Working Together Apart: Collaboration over the Internet. *Synthesis Lectures on Human-Centered Informatics* **2013**, 6 (5), 1–151.
- (43) Hall, K. L.; Vogel, A. L.; Stipelman, B.; Stokols, D.; Morgan, G.; Gehlert, S. A Four-Phase Model of Transdisciplinary Team-Based Research: Goals, Team Processes, and Strategies. *Transl. Behav. Med.* **2012**, 2 (4), 415–430.
- (44) Seering, J.; Ng, F.; Yao, Z.; Kaufman, G. Applications of Social Identity Theory to Research and Design in Computer-Supported Cooperative Work. *Proceedings of the ACM on Human-Computer Interaction* **2018**, 2, 1–34.
- (45) Kreijns, K.; Kirschner, P. A.; Jochems, W. Identifying the Pitfalls for Social Interaction in Computer-Supported Collaborative Learning Environments: A Review of the Research. *Computers in Human Behavior* **2003**, 19, 335–353.
- (46) Haas, M.; Mortensen, M. The Secrets of Great Teamwork. *Harv. Bus. Rev.* **2016**, 94 (6), 70–76.
- (47) Bai, A.; Mork, H.; Stray, V. How Agile Teams Regard and Practice Universal Design During Software Development. *Stud. Health Technol. Inform.* **2018**, 256, 171–184.
- (48) Geister, S.; Konradt, U.; Hertel, G. Effects of Process Feedback on Motivation, Satisfaction, and Performance in Virtual Teams. *Small Group Research* **2006**, 37 (5), 459–489.
- (49) Spolsky, J. Painless Functional Specifications Part 1: Why Bother? In *Joel on Software*; Apress: Berkeley, CA, 2004; pp 45–51.
- (50) Brigham, T. J. Taking Advantage of Google's Web-Based Applications and Services. *Med. Ref. Serv. Q.* **2014**, 33 (2), 202–210.
- (51) Kubaśzewski, Ł.; Kaczmarczyk, J.; Nowakowski, A. Management of Scientific Information with Google Drive. *Polym. Orthop. Traumatol.* **2013**, 78, 213–217.
- (52) Olenewa, R.; Olson, G. M.; Olson, J. S.; Russell, D. M. Now That We Can Write Simultaneously, How Do We Use That to Our Advantage? *Commun. ACM* **2017**, 60 (8), 36–43.

- (53) Vale, R. D. Accelerating Scientific Publication in Biology. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (44), 13439–13446.
- (54) Gaffney, S. G. Center for Genetically Encoded Materials (C-GEM). <http://gem-net.net> (accessed November 27, 2018).
- (55) Rivera, K.; Cooke, N. J.; Bauhs, J. A. The Effects of Emotional Icons on Remote Communication. *Conference Companion on Human Factors in Computing Systems Common Ground - CHI '96* **1996**, 99–100.
- (56) Huang, A. H.; Yen, D. C.; Zhang, X. Exploring the Potential Effects of Emoticons. *Information & Management* **2008**, *45* (7), 466–473.
- (57) Randviir, E. P.; Illingworth, S. M.; Baker, M. J.; Cude, M.; Banks, C. E. Twittering About Research: A Case Study of the World's First Twitter Poster Competition. *F1000Res* **2015**, *4*, 798.
- (58) Carrillo, F.; Cecchi, G. A.; Sigman, M.; Fernández Slezak, D. Fast Distributed Dynamics of Semantic Networks via Social Media. *Comput. Intell. Neurosci.* **2015**, *2015*, 1.
- (59) Lee, J.; Kladwang, W.; Lee, M.; Cantu, D.; Azizyan, M.; Kim, H.; Limpaecher, A.; Gaikwad, S.; Yoon, S.; Treuille, A.; et al. RNA Design Rules from a Massive Open Laboratory. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 111–122.
- (60) Rose, M. Minimal Mistakes. <https://mmistakes.github.io/minimal-mistakes/> (accessed November 29, 2018).
- (61) Gaffney, S. G. GEM-NET's RSS modification script (podcast-rss-modify). <https://github.com/gem-net/podcast-rss-modify> (accessed November 30, 2018).